STATISTICAL ANALYSIS PLAN

A Randomized, Open-label, Phase 3 Study in Subjects with Relapsed and Refractory Multiple Myeloma Receiving Carfilzomib in Combination with Dexamethasone, Comparing Once-weekly versus Twice-weekly Carfilzomib Dosing

Protocol Number: **20140355**

Version: **1.0**

Date: **09 Jun 2017**

Author: PPD

Table of Contents

List of Tables

List of Figures

List of Appendices

Table of Abbreviations

Term or Abbreviation	Description
AE(s)	Adverse Event(s)
ANC	Absolute Neutrophil Count
AUC	Area Under Curve
BSA	Body Surface Area
CBR	Clinical Benefit Rate
CDM	Clinical Data Management
CI	Confidence Interval
Cmax	Maximum plasma concentration
CR	Complete Response
CrCl	Creatinine Clearance
CSR	Clinical study report
CT-L	chymotrypsin-like
CV	coefficient of variation
BSA	Body Surface Area
CRF	Case Report Form
DMC	Data Monitoring Committee
DCB	Duration of Clinical Benefit
DOR	Duration of Response
EBMT	European Group for Blood and Marrow Transplant
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
ELISA	Enzyme-linked Immunosorbent Assay
EORTC	European Organization for Research and Treatment of Cancer
EuroQOL	European Quality of Life
EQ-5D-5L	European Quality of Life-5 Dimensions
EOI	Event of Interest
EOT	End of Treatment
E-R	Exposure-Response
FDA	Food and Drug Administration
FISH	Fluorescence in situ hybridization
GSO-DM	Global Study Operations-Data Management
HR	Hazard Ratio
IBG	Independent Biostatistics Group

IMiD	Immunomodulatory Agent
IMWG-URC	International Myeloma Working Group-Uniform Response Criteria
IPD	Important Protocol Deviation
IRC	Independent Review Committee
ISS	International Staging System
IV	Intravenous(ly)
KM	Kaplan-Meier
LC–MS/MS	Liquid chromatography–mass spectrometry/ mass spectrometry
LFT	Liver Function Test
LVEF	Left Ventricular Ejection Fraction
MedDRA	Medical Dictionary for Regulatory Activities
MR	Minimal Response
NCI-CTCAE	National Cancer Institute-Common Terminology Criteria for Adverse Events
ORR	Overall Response Rate
OS	Overall Survival
PASS	Power Analysis and Sample Size Software
PD	Progressive Disease
PDn	Pharmacodynamics
PFS	Progression-free Survival
PI	Proteasome Inhibitor
PK	Pharmacokinetics
PKDM	Pharmacokinetics and Drug Metabolism
PO	Orally
PR	Partial Response
QLQ-C30	Quality of Life Core Module
QLQ MY20	Quality of Life Multiple Myeloma Module 20
QTc	corrected QT-interval
RDI	Relative Dose Intensity
SAE(s)	Serious Adverse Event(s)
SAP	Statistical Analysis Plan
sCR	Stringent Complete Response
SD	Stable Disease
SSAP	Supplemental Statistical Analysis Plan
VGPR	Very Good Partial Response
WHODRUG	World Health Organization Drug dictionary

1. Introduction

The purpose of this Statistical Analysis Plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol for carfilzomib study 20140355, dated 08 Feb 2017. The scope of this plan includes the interim analysis and the final analysis that are planned and will be executed by the Biostatistics department unless otherwise specified. PK analyses will be provided by the Department of Clinical Pharmacology, Modeling & Simulation (CPMS).

2. Objectives

2.1 Primary Objective

The primary objective of this study is to compare the progression-free survival (PFS) of once weekly carfilzomib dosing in combination with dexamethasone to the PFS of twice weekly carfilzomib dosing in combination with dexamethasone in subjects with relapsed and refractory multiple myeloma who have received prior treatment with a proteasome inhibitor and an IMiD (immunomodulatory agent).

2.2 Secondary Objectives

The secondary objectives of the study are to compare the following between treatment groups:

- Overall response rate (ORR)
- Overall survival (OS)
- Safety and tolerability
- Pharmacokinetics (PK) of carfilzomib using sparse sampling

2.3 Exploratory Objectives

The exploratory objectives are to evaluate the following between treatment groups:

- Intensive pharmacokinetics (PK) and pharmacodynamics (PDn) of carfilzomib in a subset of subjects (substudy)
- All subscales of the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Core Module (QLQ-C30), and the EORTC Quality of Life Multiple Myeloma Module 20 (QLQ-MY20)
- European Quality of Life-5 Dimensions (EQ-5D-5L); a standardized measure of health status developed by the European Quality of Life (EuroQol) Group
- Patient reported convenience and satisfaction with the carfilzomib dosing schedule

- Healthcare resource utilization

3. Study Overview

3.1 Study Design

This is an open label, multicenter, randomized Phase 3 study comparing once weekly carfilzomib dosing in combination with dexamethasone to twice weekly carfilzomib dosing in combination with dexamethasone in subjects with relapsed and refractory multiple myeloma, 2 or 3 prior therapies, and previously treated with a proteasome inhibitor and an IMiD. The primary endpoint is PFS. The study design is illustrated in Figure 1.

Eligible subjects will be randomized in a 1:1 ratio to receive a regimen consisting of either once-weekly or twice weekly carfilzomib in combination with dexamethasone.

The randomization will be stratified by:

• International Staging System (ISS) Stage at study entry (Stage 1 versus Stage 2 or 3) per International Myeloma Working Group (Greipp 2005).

• Refractory to bortezomib treatment (Yes versus No)

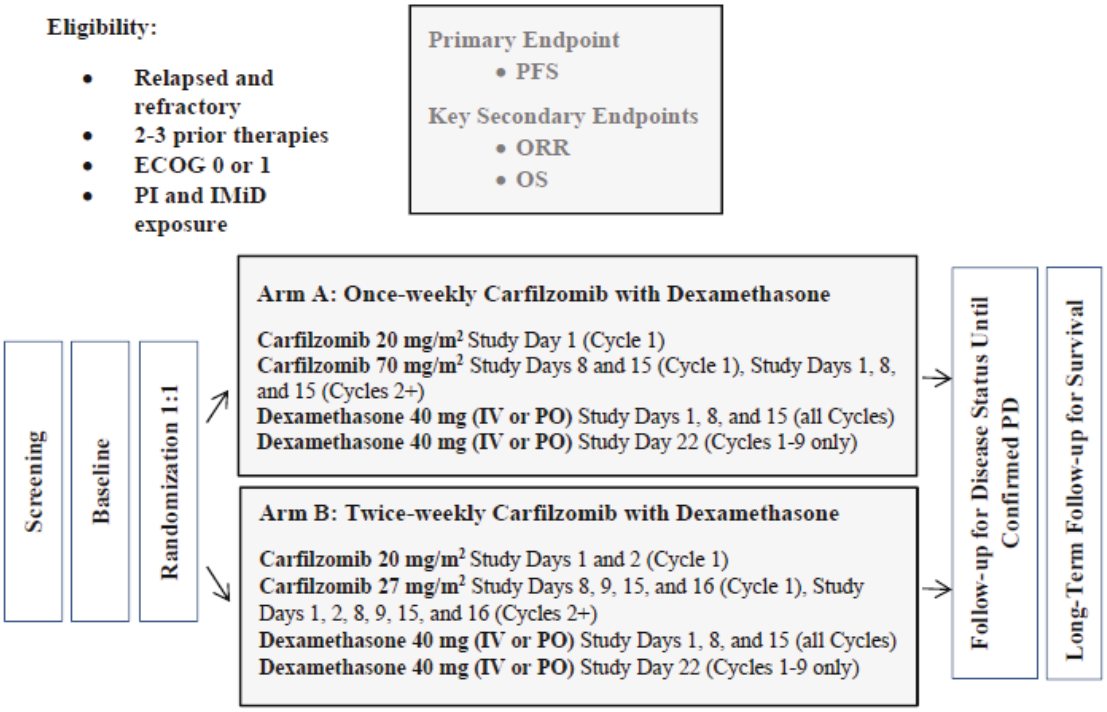
• Age (< 65 versus ≥ 65 years)

Study treatment will be administered in 28 day cycles. Subjects will receive the study treatment determined by randomization until disease progression, unacceptable toxicity, withdrawal of consent, or death (whichever occurs first). No crossover between the 2 treatment groups will be allowed.

Following termination of study treatment, all subjects will be followed for disease status (if progressive disease [PD] has not been reached), subsequent antimyeloma treatment, and survival.

All subjects will be followed for safety for at least 30 additional days after the last study treatment administration. All treatment-related adverse events (AEs) and serious adverse events (SAEs) will be followed until resolution or stabilization.

Figure 1 Study 20140355 Schema



ECOG = Eastern Cooperative Oncology Group; IMiD = immunomodulatory agents; IV = intravenous(ly); ORR = overall response rate; OS = overall survival; PD = disease progression; PFS = progression-free survival; PI = proteasome inhibitor; PO = orally.

Subjects will be assessed for response using central laboratory results every 28 days (± 4 days until confirmed PD. Long-term follow-up (LTFU) for disease status (only in cases where subjects discontinued treatment prior to PD) and for survival (after reaching PD) will continue after treatment discontinuation or until the subject has withdrawn consent for further participation, is lost to follow-up, has died, or the sponsor makes a decision to close the study. For subjects who discontinued treatment before disease progression occurred (such as for an adverse event, noncompliance, etc.), disease response assessments will be performed using central laboratory results every 28 ± 4 days until disease progression. Follow up for survival will continue approximately every 12 weeks ± 28 days, or as needed, for all surviving subjects until study closure. For any subject who is lost to follow-up, the study site will attempt to ascertain survival information via public database search.

3.2 Sample Size

CCI [redacted]

CCI [redacted]

CCI [redacted] a total of 350 PFS events will be needed for the final PFS analysis, in order to have 83% power to detect a significant difference in PFS between the two treatment groups at 1-sided significance level of 0.025 with 1 interim analysis performed at approximately 75% information time (i.e., 75% of 350 PFS events) using the Lan-DeMets spending function with an O'Brien-Fleming approach, CCI [redacted]

There is no target number of events for OS analysis. The analysis of OS will be conducted at the time of the PFS analysis (interim or final), only if both PFS and ORR analyses are positive.

4. Study Endpoints and Covariates

4.1 Study Endpoints

Primary Endpoint

The primary endpoint is PFS, which is defined as the time in months from randomization to the earlier of disease progression or death due to any cause.

Secondary Endpoints

The secondary endpoints of the study are as follows:

- ORR, defined as the proportion of subjects who achieved a confirmed PR, VGPR, CR, or sCR, according to the IMWG URC
- OS, defined as the time in months from randomization to death due to any cause
- Safety and tolerability
- Sparse PK

4.2 Planned Covariates

The primary analysis of PFS, ORR and OS will be stratified by the stratification factors at randomization: ISS stage at study entry (Stage 1 versus Stage 2 or 3), refractory to bortezomib treatment (yes versus no) , and age (< 65 versus ≥ 65 years).

5. Hypotheses and/or Estimations

The inferential comparison of PFS between treatment groups will be made using a log-rank test stratified by randomization stratification variables. The 1-sided p-value from the stratified log-rank test will be compared against the value specified by the alpha spending function specified in Section 8 at the given analysis (interim or final) to test the null hypothesis that weekly arm doesn't have superior PFS than twice weekly arm. The inferential comparison of ORR between treatment groups will be made using a Cochran-Mantel-Haenszel test stratified by randomization stratification variables. The inferential comparison of OS between treatment groups will be made using a log-rank test stratified by randomization stratification variables.

6. Definitions

Anti-myeloma therapies during long term follow-up

Anti-myeloma therapies during long term follow-up will be those therapies entered in the anti-myeloma therapy case report form (CRF) administered during the long term follow-up period of the study.

Baseline

For analyzing progression-free survival time (PFS) and overall survival time (OS), baseline will be defined as the day of randomization.

For the analysis of other endpoints, baseline will be defined as the value measured on day1 of the first cycle of protocol-specified therapy. The protocol specifies that all study procedures on day 1 should be completed before the initiation of protocol-specified therapy, which will be the assumption in the analysis, unless the time of the assessment is recorded. If a day 1 value is not available, the latest value before the day of the start of protocol-specified therapy will be used. If a subject doesn't receive any protocol-specified therapy, then the latest value prior to or on randomization date will be used.

Best overall response

Best overall response will be assessed by computational assessment, Independent Review Committee (IRC) and investigator. The primary analysis of ORR will be based on best overall response assessed by computational assessment.

Best overall response by computational assessment

Best overall response will be derived using a computer based algorithm for multiple myeloma disease assessments based on The International Myeloma Working Group Uniform Response Criteria (Durie et al. 2006, with corrections, Rajkumar et al. 2011) (IMWG-URC), the protocol, and The European Group for Blood and Marrow Transplant (EBMT) criteria (Bladé et al. 1998) for the assessment of minimal response only. The detailed algorithm is documented in a separated document 'SPECIFICATIONS FOR ONYX RESPONSE COMPUTATIONAL ASSESSMENT (ORCA) BASED ON IMWG UNIFORM RESPONSE CRITERIA'.

Best overall response by IRC assessment

Best overall response by IRC will be collected on CRF.

Best overall response by investigator assessment

Best overall response by investigator assessment is the best confirmed response by the analysis trigger date based on the responses by visit collected on myeloma response assessment eCRF. The response assessments done after initiation of new anti-myeloma therapy will be excluded from the analysis.

A confirmed response of minimal response (MR) or better requires at least two consecutive assessments with the same response or better. Stable disease (SD) requires a duration of at least 6 weeks. Progressive disease (PD) requires two consecutive PD assessments based on the same analyte except when PD is due to any of the following criteria:

- Definite development of new bone lesions or definite increase in the size of existing bone lesions
- Definite increase or new appearance of soft tissue plasmacytomas
- Definite increase of % plasma cells in Bone Marrow.

Best overall response will be decided in the following order of confirmed responses:

stringent complete response (sCR), complete response (CR), very good partial response (VGPR), partial response (PR), MR, SD, PD starting from the best to the worst.

Clinical Benefit Rate (CBR)

CBR is the proportion of subjects whose best overall response is MR, PR, VGPR, CR, or sCR.

Cumulative Dose of Study Treatment

Carfilzomib: Cumulative dose will be calculated in mg and mg/m^2. The cumulative dose in mg will be calculated as the summation of total quantity administered (mg) over all infusions. The cumulative dose in mg/m^2 will be calculated as the following over infusions.

$$\sum \frac{Total\ quantity\ administered\ (mg)}{BSA\ (m^2)}$$

BSA is to be determined by a standard formula, such as the Mosteller Formula

(Mosteller 1987): body surface area (BSA) (m^2) = ([Height (cm) × Weight (kg)]/

3600)½. The one collected on cycle 1 day 1 will be used in the calculation till when BSA changes ≥20% from cycle 1 day 1. Every time BSA changes ≥20% from current BSA used in the calculation, the new BSA will be used for subsequent infusions. If BSA is >2.2, then 2.2 will be used in the calculation.

Dexamethasone: The cumulative dose in mg will be calculated as the summation of total quantity administered (mg) over the study.

Death Date

For subjects who die during the study, the death date will be recorded on the end of study CRF in the end of study date field. Incomplete death dates where only the day of death date is missing will be imputed using the following rules:

Day 1 of the month will be used if year and month indicate that death happened in different month from last known alive date;

One day after last known alive date will be used if death happened in the same month as last known alive date.

The imputed death date will be used in calculation of duration of response, progression-free survival and overall survival.

Duration of Clinical Benefit (DCB)

DCB will be calculated only for subjects who achieve a best overall response of MR or better, i.e., sCR, CR, VGPR, PR, or MR. The duration will be calculated in months from the earliest date a response of MR or better is first achieved and subsequently confirmed to the earliest date of confirmed PD or death due to any cause:

DCB=(PD/death date-MR/better response start date +1)/30.4

Subjects will be censored using the same censoring rules for PFS as listed in table 1 if applicable.

Duration of Response (DOR)

DOR will be calculated only for subjects who achieve a best overall response of PR or better, i.e., sCR, CR, VGPR, or PR. The duration will be calculated in months from the earliest date a response of PR or better is first achieved and subsequently confirmed to the earliest date of confirmed PD or death due to any cause:

DOR=(PD/death date-response start date +1)/30.4

Subjects will be censored using the same censoring rules for PFS as listed in table 1 if applicable.

Duration of Study Treatment

Duration of treatment with carfilzomib and dexamethasone will be defined as the time from the first start date of each drug to the last stop date of the drug. Duration of the whole study treatment will be from the earliest start date among the two drugs to the latest stop date among the two drugs.

End of Study Treatment Date

The end of study treatment date is the date the decision was made to end investigational product reported on the end of investigational product administration CRF.

End of Study

For an individual subject: a subject ends the study in the event of death, , consent withdrawal, or lost to follow-up. The end of study date will be captured on the end of study CRF.

For the study as a whole: the end of study as a whole is defined as the time when the last subject is assessed or receives intervention for the purposes of final collection of data for the final analysis of PFS, which will be triggered when 350 PFS events (PD or death) are reported or by end of year 2018, whichever is earlier. Subjects who have not

completed all expected treatment at the time of the end of study as a whole will continue to follow protocol specified treatment and procedures until completion.

Enrollment Date

The date of enrollment is the date the subject gets randomized.

Fluorescent in Situ Hybridization (FISH) Risk Group

The subjects with genetic abnormalities t(4;14), t(14;16), and deletion 17p, as indicated on FISH/Conventional Cytogenetics CRF are categorized in the high risk group. The other subjects are in the standard risk group.

International Staging System (ISS) Stage at Baseline

ISS stage at baseline will be calculated using serum beta-2 microglobulin and serum albumin values collected at baseline, based on the criteria published by the International Myeloma Working Group (Greipp 2005):

Stage I: Serum beta-2 microglobulin < 3.5 mg/L and serum albumin ≥ 3.5 g/dL

Stage II:Serum beta-2 microglobulin < 3.5 mg/L and serum albumin < 3.5 g/dL or Serum beta-2 microglobulin 3.5–<5.5 mg/L irrespective of the serum albumin

Stage III: Serum beta-2 microglobulin ≥ 5.5 mg/L

Last Dose Date of Study Treatment

It is the latest date of drug taken among carfilzomib and dexamethasone.

CCI

[redacted]

[redacted]

Overall Response Rate (ORR)

ORR is the proportion of subjects whose best overall response is PR, VGPR, CR, or sCR.

Overall Survival (OS)

OS time in months will be calculated from date of randomization until death due to any cause:

$$OS=(\text{death date-randomization date } +1)/30.4$$

Subjects still alive will be censored at the date last known to be alive. If the date last known to be alive is after the date that triggers the analysis, the subject will be censored at the analysis trigger date.

Progression-Free Survival (PFS)

PFS time will be calculated from thedate of randomization until PD or death due to any cause, whichever occurs first:

$$PFS=(PD/\text{death date-randomization date } +1)/30.4$$

The duration of PFS will be right censored for subjects who meet 1 of the following conditions: 1) no baseline disease assessments; 2) starting a new anti-myeloma therapy before documentation of progressive disease or death; 3) progressive disease or death immediately after more than 1 consecutively missed disease assessment visit or; 4) alive without documentation of disease progression before the analysis trigger date. The

censoring rules for PFS primary analysis are described in Table 1. These rules are based on the May 2007 FDA Guidance for Industry, 'Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics' (www.fda.gov/cder/guidance/7478fnl.htm).

Table 1. Censoring Rules for Primary PFS Analysis

Situation	Date of Progression or Censoring	Outcome
No baseline disease assessments	Date of randomization	Censored
New anti-myeloma treatment started before documentation of PD or death	Date of last disease assessment prior to start of a new anti-myeloma treatment	Censored
Death or PD immediately after more than 1 consecutively missed disease assessment visit*	Date of last disease assessment visit before the first missed visit	Censored
Alive and without PD documentation	Date of last disease assessment	Censored
Death or PD between planned disease assessments	Date of death or first disease assessment showing PD, whichever occurs first	Progressed
Death before first disease assessment	Date of death	Progressed

* If death or PD is more than 64 days after previous disease assessment, or randomization date if there is no previous disease assessment..

Randomization Date

Randomization Date is defined as the date the subject was allocated to a treatment group. Per protocol, study treatment will ideally commence on the day of randomization, but at least within 5 calendar days of randomization.

Relative Dose Intensity (RDI)

RDI reflects whether the dose intensity of a therapy was implemented as planned. It will be calculated as the ratio of actual dose intensity relative to planned dose intensity.

$$\text{Relative Dose Intensity } (\%) = 100 \times \frac{\text{Actual Dose Intensity}}{\text{Planned Dose Intensity}}$$

Carfilzomib: Actual dose intensity is defined as the actual amount of drug in mg/m^2 delivered to a subject per week of treatment.

$$\text{Actual Dose Intensity } (\text{mg}/m^2/\text{week}) = \frac{\text{Cumulative Dose of Carfilzomib } (\text{mg}/m^2)}{\text{Number of Weeks of Actual Treatment}}$$

CCI [redacted]

CCI [redacted]

Table 2. Planned Carfilzomib Dose per Week

Arm	Cycle	Week	Protocol Specified Dose for Treatment Week (mg/m2)
Once-weekly carfilzomib with dexamethasone	1	1st	20
	2 or later	1st	70
	All cycles	2nd ,3rd	70
	All cycles	4th	0
Twice-weekly carfilzomib with dexamethasone	1	1st	40
	2 or later	1st	54
	All cycles	2nd ,3rd	54
	All cycles	4th	0

Dexamethasone: Actual dose intensity is the actual amount of drug in mg delivered to a subject per week of treatment.

$$\text{Actual Dose Intensity (mg/week)} = \frac{\text{Cumulative Dose of Dexamethasone (mg)}}{\text{Number of Weeks of Actual Treatment}}$$

CCI [redacted]

CCI

[redacted]

Refractory to prior multiple myeloma therapy

Subject is refractory to a drug received in prior regimens if the data collected on prior multiple myeloma therapy CRF indicates that any of the following criteria is met:

a. Best Response to any regimen containing the drug is stable disease or progressive disease

b. Reason the drug was stopped is progression in any regimen

c. Date of relapse/progression is after start date and within 60 days after stop date of the drug in any regimen

Study Day 1

Study day 1 is the earliest date when carfilzomib or dexamethasone is given for subjects who receive at least one dose of study drug, randomization date for subjects who are randomized but not treated.

Time to Response

Time to response will be calculated only for subjects who achieve a best overall response of PR or better, i.e., sCR, CR, VGPR, or PR. It will be calculated in months from randomization date to the earliest date a response of PR or better is first achieved and subsequently confirmed:

Time to Response=(response start date-randomization date +1)/30.4

Treatment-emergent Adverse Event

AEs starting on or after first dose of investigational product as determined by the flag indicating if the adverse event started prior to the first dose on the Events CRF and up to and including 30 days after the end of study treatment.

7. Analysis Subsets

7.1 Primary Analysis Set

The primary analysis of efficacy will be performed on all randomized subjects analyzed according to their randomized treatment assignment, regardless of the treatment received.

7.2 Safety Analysis Set

The primary analysis of safety will be performed on the Safety Analysis Set which will include all subjects who received at least one dose of study treatment analyzed according to the treatment they received.

7.3 Per Protocol Set

The Per Protocol Set will include all randomized subjects who did not have any important protocol deviations which could have an impact on the efficacy evaluation of the subject.

Subjects with the following protocol deviations will be excluded from the per protocol set for OS analysis:

- Inclusion criteria of relapsed multiple myeloma
- Inclusion criteria of refractory multiple myeloma
- Inclusion criteria of 2 or 3 prior therapies for multiple myeloma
- Inclusion criteria of prior exposure to an IMiD
- Inclusion criteria of prior exposure to a Proteasome Inhibitor (PI)
- Inclusion criteria of documented response of at least partial response (PR) to at least 1 prior line of therapy.
- Inclusion criteria of measurable disease
- Exclusion criteria of second malignancy within the past 5 years
- Exclusion criteria of cytotoxic chemotherapy or other antineoplastic within 28 days prior to randomization
- Exclusion criteria of immunotherapy such as an IMiD or PI within 21 days prior to randomization

- Exclusion criteria of glucocorticoid therapy exceeding a cumulative dose of 160mg within 21 days prior to randomization
- Subject received concurrent therapy with a marketed or investigational anticancer therapeutic or radiation to large marrow reserves for either a therapeutic or palliative intent while receiving study treatment
- Subject received wrong treatment other than treatment assigned to

In addition to the deviations listed above, subjects who had the following deviations will also be excluded from per-protocol set for PFS and ORR analysis:

- Subject permanently discontinued treatment due to progressive disease (PD) based on local labs not central labs
- Subject ended treatment before confirmed PD but no further disease assessments were carried out
- Baseline laboratory disease assessment was not done in such a way subject is not measurable as defined per protocol
- Subject received plasmapheresis while receiving study treatment

Subjects will be analyzed according to their randomized treatment assignment.

7.4 Health-related Quality of Life Analyses Set

All randomized subjects will be included in Health-related Quality of Life Analyses Set.

7.5 Pharmacokinetic/Pharmacodynamic Analyses Set

All subjects who participated in the intensive PK/PDn substudy or the sparse PK sampling will be included in the Pharmacokinetic/Pharmacodynamic Analyses Set. These subjects will be evaluated for pharmacokinetics and pharmacodynamics unless significant protocol deviations affect the data analysis or if key dosing, dosing interruption, or sampling information is missing.

7.6 Interim Analyses Set

The interim analyses set will include all randomized subjects. The independent safety reviews, which are scheduled to occur approximately every 6 months, will include all randomized subjects for efficacy analyses and all treated subjects at the time of the database snapshot for a given review.

7.7 Subgroup Analyses

Exploratory subgroup analyses for the primary endpoint and key secondary endpoints (ORR and OS) will be performed for each level of the three randomization stratification factors: International Staging System (ISS) Stage at study entry (Stage 1 vs. Stage 2 or 3), Refractory to bortezomib treatment (Yes vs. No) and Age (< 65 vs. ≥ 65 years, <75 vs. ≥ 75). Additional subgroup analyses will be based on the following factors:

- Sex (male vs. female)
- Race (White vs. Asian vs. Other)
- Ethnicity (Hispanic or Latino vs. Not Hispanic or Latino)
- Geographic region (Europe vs. North America vs. Asian Pacific)
- Baseline ECOG (0 vs. 1)
- Baseline creatinine clearance (30 - < 50 ml/min vs. 50 - < 80 ml/min vs. ≥ 80 ml/min)
- FISH risk group (high risk vs. standard risk vs. unknown)
- β2 microglobulin level (< 3.5 mg/L vs. ≥ 3.5 mg/L)
- ISS stage at baseline (Stage I vs. Stage II and III)
- ISS stage at diagnosis (Stage I vs. Stage II and III vs. Unknown)
- Corrected calcium at baseline (≤11.5 mg/dL vs. >11.5 mg/dL)
- Lines of prior treatment (2 vs. 3)
- Prior transplant (Yes vs. No)
- Prior lenalidomide (Yes vs. No)
- Refractory to lenalidomide (Yes vs. No)

These subgroups, except for stratification variables, will be re-examined and appropriately re-categorized before unblinding. Subgroup analysis containing a subgroup of less than 5% of the whole population will not be conducted.

8. Interim Analysis and Early Stopping Guidelines

One interim analysis will be conducted for PFS when approximately 263 (75% of total 350) PFS events are observed. The objective of the interim analysis is to monitor for differences between treatment arms for evidence of a substantial benefit in the carfilzomib once-weekly with dexamethasone arm.

To ensure proper control of type I error rate, the interim and final analysis of PFS will be analyzed under a group sequential design framework with the stopping boundaries constructed using the Lan-DeMets spending function with an O'Brien-Fleming approach. The stopping boundaries in an example scenario are calculated using East 6.4 and presented in Table 3. Actual boundaries will be calculated on the basis of observed PFS events up to the data cut-off date for the interim analysis. .

CCI

An external independent Data Monitoring Committee (DMC) will review the interim PFS analysis results and recommend whether the study should stop early for efficacy. If the DMC recommends to stop the study for efficacy, the sponsor will consider stopping the trial and conclude that the carfilzomib once-weekly with dexamethasone arm prolongs PFS compared to the carfilzomib twice-weekly with dexamethasone arm. In addition, the DMC will assess safety approximately every 6 months. An Independent Biostatistics Group (IBG) will perform the analyses and provide the interim reports to the DMC. The IBG and DMC will have access to subjects' individual treatment assignments.To minimize the potential introduction of bias to the conduct of the study, members of the DMC and IBG will not have any direct contact with study center personnel or subjects. The DMC will communicate major safety concerns and recommendations regarding study modification or termination based on the safety and efficacy parameters to Amgen in accordance with the DMC charter.

Records of all meetings will be maintained by the DMC for the duration of the study. Records of all meetings will be stored in the Amgen official document management system at the conclusion of the study. Further details are provided in the DMC charter.

9. Data Screening and Acceptance

9.1 General Principles

The objective of the data screening is to assess the quantity, quality, and statistical characteristics of the data relative to the requirements of the planned analyses. The database will be subject to edit checks outlined in the data management plan by Amgen Clinical Data Management (CDM) department.

9.2 Data Handling and Electronic Transfer of Data

The Amgen Global Study Operations-Data Management (GSO-DM) department will provide all data to be used in the planned analyses. This study will use the RAVE database. Laboratory data will be collected by COVANCE Central Laboratory Services and transferred to Amgen GSO-DM weekly in cumulative files. Quality of life data will be collected by ERT and transferred to Amgen GSO-DM monthly in cumulative files.

9.3 Handling of Missing and Incomplete Data

The descriptive statistics will identify the extent of missing data. Rules for handling missing data related to endpoints are described in the endpoint definitions (Section 6) or in the description of analyses (Section 10). The handling of incomplete and partial dates for adverse events and concomitant medications are described in Appendix B. Handling of missing or incomplete data for Exposure-Response (E-R) analysis will be described in the documents to support population PK/PDn dataset generation and E-R analysis.

9.4 Detection of Bias

Methods to detect bias are described in the analyses of particular endpoints (Section 10).

9.5 Outliers

Any suspected outliers will be investigated by the study team and will be included in the database unless determined to be an error or there is supporting evidence or explanation to justify the exclusion. Any outliers excluded from the analysis will be discussed in the Clinical Study Report (CSR), including the reasons for exclusion and the impact of their exclusion on the study. PK concentration data will be evaluated for outliers by visual inspection, and decisions to re-assay individual samples will be made in accordance with standard PKDM practice.

9.6 Distributional Characteristics

The statistical assumptions for analysis methods will be assessed. If the assumptions for the distributional characteristics are not met, these will be described and further analyses may be carried out using data transformations or alternative analysis methods. The use of transformations or alternative analysis methods will be justified in the final study report.

9.7 Validation of Statistical Analyses

Programs will be developed and maintained, and output will be verified in accordance with current risk-based quality control procedures.

Tables, figures and listings will be produced with validated standard macro programs where standard macros can produce the specified outputs.

The production environment for statistical analyses consists of Amgen-supported versions of statistical analysis software, for example the SAS System version 9.3 or later. For the PK/PDn analysis, refer to section 10.7 for software used.

10. Statistical Methods of Analysis

10.1 General Principles

Continuous variables will be summarized by the non-missing sample size (n), mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized by the n and percentage in each category. Time to event endpoints will be summarized with Kaplan-Meier (KM) curves, KM proportions at select time points, KM quartiles (when estimable), the number of subjects with events, the number of subjects censored, and censoring reasons. Duration of follow up for time to event endpoints will be estimated using the reverse Kaplan Meier method (Schemper and Smith 1996). Point estimates for efficacy endpoints will be accompanied by 2-sided 95% confidence intervals including estimates of KM quartiles (Klein and Moeschberger, 1997), KM proportions (Kalbfleisch and Prentice, 1980), and binomial proportions (Clopper CJ and Pearson, 1934). To examine non-proportionality between the treatment groups for time to event endpoints, the log cumulative hazard curves will be plotted. The scaled Schoenfeld residuals by time plot will be examined for evidence of a non-zero correlation, which indicates non-proportionality. Interaction between treatment and the log event time will be also tested at 10% level in a Cox model stratified by randomization stratification variables to test for non-proportionality

The hypotheses for the primary efficacy endpoint, PFS, and secondary efficacy endpoints (ORR and OS) will be tested using a fixed sequence hierarchical testing procedure to control the family-wise Type I Error rate below 1 sided 0.025 level. The family of hypotheses is ordered as follows: 1) PFS, 2) ORR, and 3) OS. Starting with the hypothesis of PFS, if any hypothesis in the sequence is rejected, then the subsequent hypothesis will be tested; if any hypothesis is accepted, then the subsequent hypotheses will not be tested. Furthermore, potential multiplicity from testing at two possible times of PFS IA and final analysis for ORR and OS will be adjusted using the Lan-DeMets spending function with an O'Brien-Fleming approach and a Pocock approach respectively (Hung et al 2007). CCI [redacted]

CCI

[redacted]

10.2 Subject Accountability

The number and percent of subjects who were screened, randomized, received study treatment, entered long-term follow-up before disease progression and long-term follow-up for survival will be summarized by treatment group. The number and percent of subjects who discontinued study treatment, long-term follow-up before disease progression, and study will be tabulated, along with the reason for discontinuation. The number and percent of subjects randomized will be tabulated by the stratification factors. The number and percent of subjects randomized will be tabulated by study site. Key study dates for the first subject randomized, last subject randomized, and data cut-off date for analysis will be presented.

10.3 Important Protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first subject's visit and updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, sub-category codes, and descriptions will be used during the course of the study. Eligibility deviations are defined in the protocol.

10.4 Demographic and Baseline Characteristics

Demographic (ie, age, age group [<65, 65-74, 75-84, ≥85], sex, race, and ethnicity) and baseline disease characteristics will be summarized by treatment group and overall using descriptive statistics for the Primary Analysis Set. The baseline characteristics to be summarized include:

- Geographic region (Europe, North America, Asian Pacific)
- Reproductive status
- Baseline ECOG
- Baseline laboratories including: hemoglobin, ANC, platelet count, and creatinine clearance
- Baseline LVEF (%)
- Time from initial multiple myeloma diagnosis to randomization (months)
- Measurable disease category
- M-protein heavy and light chain subtype
- FISH risk group
- Serum free light chain Kappa/Lambda ratio
- Presence of any plasmacytoma
- Presence of any bone lesion
- Corrected calcium at baseline
- β2 microglobulin level
- ISS stage at diagnosis
- ISS stage at baseline
- Lines of prior treatment
- Prior transplant
- Prior bortezomib
- Prior lenalidomide
- Prior thalidomide
- Refractory to bortezomib
- Refractory to lenalidomide

10.5 Efficacy Analyses

10.5.1 Analyses of Primary Efficacy Endpoint

1-sided stratified log-rank test, stratified by the randomization factors, will be used to compare PFS between treatment groups as primary inference. In addition, a hazard ratio with a 95% confidence interval will be estimated from a stratified Cox regression model. The KM summaries described in Section 10.1 will be performed by treatment group. The primary analysis will be performed on the Primary Analysis Set based on computational assessments. Sensitivity analyses will be performed using investigator assessments, Independent Review Committee (IRC) assessments, different censoring rules regarding use of new anti-myeloma therapy and different analysis population. The details are documented in table 5. The discordance between the results regarding progression status and timing from the internal computational assessment, investigator and IRC will be summarized. Subgroup analyses will be performed to explore the consistency of the treatment effect for subgroups described in Section 7.7. Piecewise Cox models for 3-month time intervals may be explored given evidence of non-proportional hazards

(Collett, 2003). This model will allow estimation of an overall weighted hazard ratio (weights equal to fraction of total events in each interval (Lu & Pajak, 2000)) as well as within interval treatment hazard ratio. Additional analysis may be performed to explore potential sources for non-proportionality by considering baseline prognostic factors and other potential confounding factors.

10.5.2 Analyses of Secondary Efficacy Endpoints

The ORR will be analyzed at the time of the PFS analysis (interim or final), only if the PFS analysis crosses the stopping boundary. The ORR will be calculated by treatment group and the associated 95% CI will be estimated using the Clopper Pearson method. The inferential comparison of ORR between treatment groups will be performed using Cochran-Mantel-Haenszel test stratified by the randomization stratification factors. 1-sided p-value from the test will be compared against the alpha value specified in Table 4 to determine the significance. Odds ratio (and its 95% CI) will be estimated using Mantel-Haenszel method . The primary analysis of ORR will be based on computational assessments for the Primary Analysis Set. Sensitivity analyses will be performed based on investigator assessments, Independent Review Committee (IRC) assessments, unstratified method, and for Per Protocol Set. The discordance between the results from the internal computational assessment, investigator and IRC will be summarized. Subgroup analyses will be performed to explore the consistency of the treatment effect for subgroups described in Section 7.7.

The analysis of OS will be conducted at the time of the PFS analysis (interim or final), only if both PFS and ORR analyses are positive. The same method as for analysis of PFS will be used for analysis of OS.The primary analysis of OS will be performed on the Primary Analysis Set. Sensitivity analyses will be performed based on unstratified analyses, and for Per Protocol Set. Subgroup analyses will be performed to explore the consistency of the treatment effect for subgroups described in Section 7.7. Piecewise Cox models for 6-month time intervals may be explored given evidence of non-proportional hazards.

Table 5. Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
Primary Endpoint		
Progression-Free Survival	Based on internal computational assessments: • KM summaries • 1-sided p-value from stratified log-rank test • Hazard ratio and 95% CI from stratified Cox regression	• Investigator assessments: same as primary summary and analysis method based on investigator assessments. • IRC assessments: Same as primary summary and analysis method based on IRC assessments. • Unstratified analyses: 1-sided p-value from unstratified log-rank test, hazard ratio and 95% CI from unstratified Cox regression. • Initiation of new anti-myeloma therapy treated as PFS Event: The data censoring rules are the same as those for the primary analysis of PFS except that the use of new anti-myeloma therapy will be treated as an event rather than a mechanism for censoring. The same analysis method as for primary analysis will be used. • Initiation of new anti-myeloma therapy treated as neither a PFS event nor a censoring event: The data censoring rules are the same as those for the primary analysis of PFS except that the initiation of new anti-cancer therapy will be excluded as a mechanism for censoring. The same analysis method as for primary analysis will be used. • Per-Protocol subset: same as primary summary and analysis method for Per-Protocol subset. • Lost to follow-up or withdrawal of consent treated as event: The data censoring rules are the same as those for the primary analysis of PFS except that subjects who were lost to follow-up or withdrew consent without PD/ death are treated

		as having an event at the next scheduled assessment time in both treatment arms. The same analysis method as for primary analysis will be used.
Secondary Endpoints		
Overall Response Rate	Based on internal computational assessments: • Point estimate of ORR and 95% CI by treatment group using the Clopper Pearson method • 1-sided p-value from Cochran-Mantel - Haenszel test stratified byrandomization stratification factors • Odds ratio associated with 95% CI using Mantel-Haenszel method	• Investigator assessments: Same as primary summary and analysis method based on investigator assessments. • IRC assessments: Same as primary summary and analysis method based on IRC assessments. • Unstratified analyses: 1-sided p-value is from Fisher exact test and odds ratio (95%CI) estimate is not adjusted for randomization stratification factors. • Per-Protocol subset: Same as primary summary and analysis method for Per-Protocol subset.
Overall Survival	• KM summaries • 1-sided p-value from stratified log-rank test • Hazard ratio and 95% CI from stratified Cox regression	• Unstratified analyses: 1-sided p-value from unstratified log-rank test, hazard ratio and 95% CI from unstratified Cox regression. • Per-Protocol subset: same as primary summary and analysis method for Per-Protocol subset.

10.5.3 Exploratory Efficacy Analysis

The KM summaries described in Section 10.1 will be performed by treatment group for duration of response and clinical benefit.

Time to response will be summarized by the non-missing sample size (n), mean, standard deviation, median, minimum, and maximum for responders by treatment group.

QLQ-C30, QLQ-MY20 and EQ-5D-5L questionnaires will be administered prior to study treatment on Day 1 of Cycle 1, then every second cycle (Cycle 1, 3, 5, etc) during treatment and every 12 weeks (every 84 days ± 4 days) until progression, or withdrawal of consent during LTFU. Patient convenience and satisfaction questionnaire will be collected at Cycle 2 and EOT only. All QLQ-C30 and QLQ-MY20 subscale scores will

be summarized descriptively by visit and by treatment group using the number of subjects with non-missing data (n), mean, standard deviation, median, minimum and maximum. Change from baseline will be summarized using the same statistics. EQ-5D-5L data will be summarized by presenting the frequency and proportion of each level for each dimension by visit and by treatment group. EQ VAS data arm will be summarized using the same method for QLQ-C30 and QLQ-MY20 subscale scores. Patient convenience and satisfaction questionnaire has only two questions about whether the patients feel the carfilzomib dosing schedule convenient and how statisfied/dissatisfied they are with the dosing schedule. The frequency and proportion of each category will be summarized for the two questions by treatment group at cycle 2 and EOT. The comparison analyses of health-related quality of life between treatment groups will be implemented in addition to the descriptive analyses described above. For instance, Global Health Status/QoL scale, an important subscale of QLQ-C30, will be compared between treatment groups across visits using a restricted maximum likelihood-based mixed model for repeated measures (MMRM) under the assumption of missing at random (MAR) (Mallinckrodt et al, 2008).

10.6 Safety Analyses

10.6.1 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 19.1 or later will be used to code all adverse events (AE) to a system organ class and a preferred term. AEs of interest (EOI) categories will be based on search strategies defined by the EOI steering committee. All adverse event tables will be summarized by treatment group.

The subject incidence of AEs will be summarized for all treatment-emergent AEs, ≥ grade 3 TEAEs, serious AEs, AEs leading to modification (i.e., dose reduction or dose interruption) of investigational product, AEs leading to withdrawal of investigational product, and fatal AEs.

Subject incidence of all treatment-emergent AEs, ≥grade 3 TEAEs ,serious AEs, AEs leading to modification (i.e., dose reduction or dose interruption) of investigational product, AEs leading to withdrawal of investigational product, and fatal AEs will be tabulated by system organ class and preferred term in descending order of frequency; similar summaries will be repeated for EOIs. Time to onset and duration of select EOIs may also be summarized.

Summaries of treatment-emergent AEs and serious AEs will be presented by preferred term in descending order.

A summary of treatment-emergent AEs will be tabulated by system organ class, preferred term, and causal relationship to study drug.

A summary of treatment-emergent AEs will be tabulated by system organ class, preferred term, and worst grade based on NCI-CTCAE Version 4.03.

Subgroup analyses (if there is a medical rationale) will be presented by system organ class and preferred term or EOI in descending order of frequency.

10.6.2 Laboratory Test Results

For hematology, chemistry, and other laboratory values, the baseline values and, changes from baseline by visit, the minimum, maximum, and last observed values will be summarized descriptively.

For the summary of changes from baseline by visit, subjects without a baseline and/or post baseline value will be excluded; values from unscheduled assessments will be excluded. Subjects with missing data for a scheduled assessment time point will be excluded from the summary for that time point. Laboratory results from samples taken > 30 days after the last administration of protocol therapy will be excluded from the laboratory summaries.

Laboratory test results will be graded using the NCI CTCAE (Version 4.03). Shifts in laboratory toxicity grades to outside the normal range will be evaluated for selected laboratory parameters defined in Appendix C by assessing the maximum increase and/or decrease observed during the course of study treatment relative to the baseline toxicity grade.

The subject incidence of Grade 3 and 4 hematological laboratory abnormalities (including neutropenia, thrombocytopenia, and anemia) will be provided by treatment group.

The subject incidence of Grade 3 and 4 nonhematological toxicities (including liver function test [LFT], CrCl) will be provided by treatment group.

10.6.3 Vital Signs

Vital sign results (systolic and diastolic blood pressure, pulse, respiratory rate, and temperature) will be summarized descriptively for baseline values, changes from baseline by visit, the minimum, maximum, and last observed values.

For the summary of changes from baseline by visit, subjects without a baseline and/or post baseline value will be excluded; values from unscheduled assessments will be

excluded. Vital sign results taken > 30 days after the last administration of protocol therapy will be excluded from all vital sign summaries.

Shifts in scores for ECOG performance status scores between baseline and end of treatment visit will be tabulated by treatment group.

10.6.4 Electrocardiogram (ECG)

Twelve lead electrocardiograms (ECGs) including corrected QT-interval (QTc; representing the corrected duration of ventricular electrical activity) will be performed locally. Electrocardiograms will be required from all subjects at Screening and EOT. ECG results (PR, QRS, QT, QTc, heart rate, and result) will be summarized descriptively by treatment group.

10.6.5 Exposure to Investigational Product

Descriptive statistics will be produced to describe the exposure to investigational product by treatment group for subjects in the Safety Analysis Set. The number of cycles of protocol-specified therapy administered will be summarized with an additional breakdown of the number of cycles started. In addition, the duration of therapy, the cumulative dose, and the average dose per administration and relative dose intenstiy will be summarized for each drug. The number and percent of subjects with dose modifications (eg, dose reductions, dose interruptions) and reason for modification will be summarized for both treatment groups.

10.6.6 Exposure to Concomitant Medication

The number and proportion of subjects receiving concomitant medications from study day 1 through safety follow-up will be summarized by preferred term as coded by the World Health Organization Drug (WHODRUG) dictionary by treatment group in the Safety Analysis Set. In addition, the number and proportion of subjects receiving anti-myeloma therapies during long term follow-up will be summarized by WHODRUG preferred term for each treatment group in the Full Analysis Set.

10.7 Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Analysis

Analysis plan for sparse PK data

Blood samples will be collected on Day 1 of Cycle 2 at specified times, processed and stored until analysis.

Plasma samples will be analyzed by LC–MS/MS analysis to measure the plasma concentration of carfilzomib. Sparse PK data collected from this study will be used to

update a population PK model previously developed (TR-1162-171) and will be described in a separate report.

The analysis population will contain all subjects who have at least one PK sample collected according to the protocol and laboratory manual. Actual collection times will be recorded and used in the analysis. The population modeling program, NONMEM (version 7.2 or higher).will be used to fit a nonlinear mixed effects model to estimate PK parameters including clearance and volume of distribution, the inter- and intra-subject variability and the population variability in the parameter estimates. The PK concentrations obtained from subjects who participate in the sparse PK sampling along with results from the intensive PK/PDn sub-study and other carfilzomib studies will be used in the development of a structural model. The best model will be evaluated by goodness-of-fit statistics and reduction in the objective function and posterior predictive checks. Subject characteristics such as age, gender, body weight, BSA, and race will be included in the model to identify potential covariates affecting PK of carfilzomib.

Analysis plan for PK and PDn sub-study

Intensive pharmacokinetic (PK) and pharmacodynamic (PDn) samples will be obtained as a sub-study (approximately n = 15 in each arm) from subjects who consent to the additional testing at selected sites. Pharmacokinetic samples will be collected on Day 15 of Cycle 1 and PDn samples will be collected on Days 1, 5, 6, and 7 of Cycle 1 and day 2 of Cycle 2.

For both arms, PK plasma samples will be collected from all subjects for determination of plasma concentrations of carfilzomib on treatment Day 15 of Cycle 1 at the following time points:

- Predose (within 5 minutes before the start of infusion)
- 15 minutes (± 5 minutes) after the start of infusion for Arm A only (once-weekly carfilzomib)
- Immediately prior to (within 2 minutes before) the end of infusion
- 15 minutes (± 5 minutes) after the end of infusion
- 60 minutes (± 5 minutes) after the end of infusion
- 2 hours (± 5 minutes) after the end of infusion

PK

The analysis population will contain all subjects who have at least one PK sample collected according to the protocol and laboratory manual. Actual collection times will be

recorded and used in the analysis. Individual and mean plasma concentration versus time data will be tabulated and plotted by dose level. The PK parameter will be estimated based on noncompartmental methods. The PK parameter estimates for carfilzomib will be summarized, including total plasma exposure (AUC, C_{max}, time to maximum plasma concentration [t_{max}], total plasma clearance, and $t_{1/2}$ [as appropriate for data collected]). Estimates for these parameters will be tabulated and summarized (i.e., n, Mean, Standard deviation, CV%, Median, Min, and Max, GeoMean and GeoCV%). These estimates will be summarized descriptively by each arm (Once-weekly vs Twice-weekly). Plasma PK parameters of carfilzomib will be computed in WinNonlin® Enterprise v.5.2 or higher.

PDn

PD analyses will be performed using a whole-blood 20S proteasome specific activity inhibition assay. The percent inhibition of 20S proteasome activity relative to baseline will be determined by a fluorogenic and/or luminescent substrate assay for the chymotrypsin-like (CT-L) (Beta5 subunit of the constitutive proteasome and LMP7 subunit of the immunoproteasome), the trypsin-like (T-L), and caspase-like (C-L) activities of the proteasome. Individual and the mean percent inhibition of 20S proteasome activity relative to baseline will be determined and will be tabulated and plotted versus nominal sampling time by each arm (Once-weekly vs Twice- weekly).

10.7.1 Exposure-Response (E-R) Analysis

PK data collected from individuals who consent to the sparse PK sampling or PK/PDn sub-study will be used to explore the E-R relationship for efficacy and safety endpoints. The data from this study will be used to update a previously developed E-R model (Amgen Pharmacometrics Report: 121604) and will be described in a separate report.

In b*r*ief, each exposure metric will be calculated using the individual subject's parameters from the population PK model (see Section 10.7 Analysis plan for sparse PK data) and dosing history. The exposure metrics will be computed from the estimated Cmax and AUC and combined with the response/adverse event data to examine the E-R relationships for efficacy and safety endpoints. A pooled analysis of data from this study and other carfilzomib clinical studies will be conducted. In addition, an E-R analysis based on this study alone may also be conducted.

Categorical responses will be tabulated by count and proportion by study. In the analysis, a quantitative assessment of the effect of each of the exposure metrics will be

performed either using logistic regression for binary endpoints or log-rank test by exposure quartile for time-to-event endpoints. Endpoints exhibiting a statistically significant relationship to carfilzomib exposure will be further evaluated with multiple logistic or Cox proportional hazards models (for time-to-event endpoints) to assess the effects of covariates in conjunction with carfilzomib exposure on efficacy and safety responses. Additional multiple logistic regression analysis will not be performed on the endpoints that showed inverse relationship.

Efficacy endpoints

The efficacy endpoints will include Progression-free survival (PFS), ORR and CBR.

Safety endpoints

Safety endpoints will include the subject incidence of any grade adverse events leading to study drug discontinuation, any Grade 3 or higher adverse events , serious adverse events, TEAEs with a fatal outcome, and cardiac, hepatic, and renal adverse events. Additional endpoints of interests may be included in the analysis.

Covariates of interest

Endpoints exhibiting a statistically significant relationship to carfilzomib exposure will be further evaluated with multiple logistic models. The key covariates tested in the multiple logistic regression analysis may include the following:

bortezomib refractory in any prior regimen (yes/no), number of prior regimens, Eastern Cooperative Oncology Group (ECOG) performance status, sex, age, race, baseline CrCL, baseline hemoglobin (g/L), baseline serum beta-2 microglobulin (mg/L), and baseline platelet count (<150, ≥150 [10^9/L]). Additional covariates of interests may be included in the analysis.

11. Changes From Protocol-specified Analyses

In the protocol, it specifies that logistic regression model will be used for ORR analysis. However, since Cochran-Mantel -Haenszel test is more widely used in analysis of binary endpoints to support regulatory filings, Cochran-Mantel-Haenszel test is specified to be used for ORR analysis in the SAP. The CMH test statistic and the score test statistic of the conditional logistic regression are approximately identical (Day N.E.& Byar D.P., 1979), so this analysis method change will not affect the conclusion of ORR analysis.

12. Literature Citations / References

Bladé J, Samson D, Reece D, et al. European Group for Blood and Marrow Transplant. Myeloma Subcommittee of the EBMT. Criteria for evaluating disease response and progression in subjects with multiple myeloma treated by high-dose therapy and haemopoietic stem cell transplantation. Br J Haematol. 1998 Sep; 102(5):1115–23.

Clopper CJ and Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika. 1934; 26(4):404-413.

Collett D. Modelling Survival Data in Medical Research. 2nd edition. London, UK: Chapman & Hall/CRC; 2003.

Day N.E., Byar D.P. (September 1979), Testing hypotheses in case-control studies-equivalence of Mantel–Haenszel statistics and logit score tests, Biometrics: 623–630

Durie BG, Harousseau JL, Miguel JS, et al. International Myeloma Working Group. International uniform response criteria for multiple myeloma. Leukemia. 2006; 20(9):1467–73. Erratum in: Leukemia 2006; 20(12):2220. Leukemia 2007; 21(5):1134.

Greipp PR, San Miguel J, Durie BG, et al. International staging system for multiple myeloma. J Clin Oncol. 2005;23(15):3412–20.

Hryniuk W, Goodyear, M. The calculation of received dose intensity. Journal of Clinical Oncology 8:1935–1937, 1990

Hung, H.M.,Wang, S. J. and O'Neil, R. (2007). Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. Journal of Biopharmaceutical Statistics 17, 1201–1210.

Kalbfleisch, J. D. and Prentice, R. L. The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons; 1980

Klein, J. P. and Moeschberger, M. L. (1997), Survival Analysis: Techniques for Censored and Truncated Data, New York: Springer-Verlag.Longo D, Duffey P, DeVita V, Wesley M, Hubbard S, Young R. The calculation of actual or received dose intensity: A comparison of published methods. Journal of Clinical Oncology 9:2042–2051, 1991

Lu, J., and Pajak, T. F. (2000), Statistical Power for a Long-Term Survival Trial with a Time-Dependent Treatment Effect, Controlled Clinical Trials, 21, 561–573.

Mallinckrodt CH, Lane PW, Schnell D, Peng Y, and Mancuso, JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Information Journal, Vol. 42, pp.303-319, 2008

Mosteller RD. Simplified calculation of body surface area. N Engl J Med 1987;317(17):1098 (letter).

Rajkumar SV, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the international myeloma workshop consensus panel 1. Blood 117:4691-4695.2011

Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Controlled Clinical Trials 17:343–346, 1996

US Food & Drug Administration: Guidance for Industry. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. www.fda.gov/cder/guidance/7478fnl.htm

13. Appendices

CCI

[redacted]

Appendix B. Reference Values/Toxicity Grades

Laboratory Values

Safety laboratory values below a distinct limit (eg. detection limit, documented as "< [limit]") will be substituted by half of the limit and values above a distinct limit (documented as "> [limit]") will be substituted by the limit itself for all analyses.

A Grade (based on CTC AE version 4.0 [v4.03: June 14, 2010]) will be assigned to each laboratory result as detailed in Table 2. Depending on the toxicity definition, the same result may be assigned to two grading for deviations towards higher or lower values. Values not meeting any of the criteria will be assigned a grade 0.

Table 2. Grading of Select Laboratory Parameters

Laboratory Abnormality [Unit]	Grade 1	Grade 2	Grade 3	Grade 4
Decreased Lymphocytes [G/L]	0.8 - < LLN	0.5 - < 0.8	0.2 - < 0.5	< 0.2
Decreased Neutrophils [G/L]	1.5 - < LLN	1.0 - < 1.5	0.5 - < 1.0	< 0.5
Decreased Platelets [G/L]	75 - < LLN	50 - < 75	25 - < 50	< 25
Decreased WBC [G/L]	3.0 - < LLN	2.0 - < 3.0	1.0 - < 2.0	< 1.0
Decreased Hemoglobin [g/L]	100 - < LLN	80 - < 100	< 80	not defined
Decreased Albumin [g/L]	30 - < LLN	20 - < 30	< 20	not defined
Increased AST	> ULN – 3*ULN	> 3*ULN – 5*ULN	> 5*ULN – 20*ULN	> 20*ULN
Increased ALT	> ULN – 3*ULN	> 3*ULN – 5*ULN	> 5*ULN – 20*ULN	> 20*ULN
Increased Total Bilirubin	> ULN – 1.5*ULN	>1.5*ULN – 3*ULN	> 3*ULN – 10*ULN	> 10*ULN
Increased Creatinine	>1 - 1.5 * BL or >ULN - 1.5 *ULN	>1.5 - 3.0 * BL or >1.5 - 3.0 * ULN	>3.0 * BL or >3.0 - 6.0 * ULN	>6.0 * ULN
Decreased Corrected Calcium [mmol/L]	2.0 - < LLN	1.75 - < 2.0	1.5 - < 1.75	< 1.5
Increased Corrected Calcium [mmol/L]	>ULN - 2.9	>2.9 – 3.1	>3.1 – 3.4	>3.4
Decreased Potassium [mmol/L]*	3.0 - < LLN	3.0 - < LLN; symptomatic; intervention indicated	2.5 - < 3.0	< 2.5

Laboratory Abnormality [Unit]	Grade 1	Grade 2	Grade 3	Grade 4
Increased Potassium [mmol/L]	5.5 - >ULN	>5.5 – 6.0	>6.0 – 7.0	>7.0
Decreased Sodium[mmol/L]	130-<LLN	not defined	120-<130	<120
Increased Sodium[mmol/L]	>ULN-150	>150-155	>155-160	>160
Decreased Phosphorus [mmol/L]	0.8 - <LLN	0.6 - <0.8	0.3 - <0.6	<0.3
Increased Magnesium [mmol/L]	>ULN – 1.23	not defined	>1.23 – 3.30	>3.30
Decreased Magnesium [mmol/L]	0.5 - <LLN	0.4 - <0.5	0.3 - <0.4	<0.3
Increased Uric Acid [mmol/L]^	>ULN – 0.59 without physiologic consequences	not defined	>ULN – 0.59 with physiologic consequences	>0.59

BL: baseline value, LLN: Lower limit of normal, ULN: Upper limit of normal
Clinical criteria from CTCAE 4.0 grading were not considered in order to assign grades unless specified otherwise

*: Details will be recorded on Potassium CTC Details CRF for central lab values and Chemistry (Local Lab) CRF for local lab values.

^: Details will be recorded on Uric Acid CTC Details CRF for central lab values and Chemistry (Local Lab) CRF for local lab values.

Appendix C. Patient-reported Outcome Forms/Instruments

The following sections describe the scoring algorithms used for both the QLQ-C30 and QLQ-MY20 questionnaires. Scoring procedures are similar for both questionnaires and can be found in the EORTC QLQ-C30 Scoring Manual, ver. 3 (Fayers et al. 2001) and Cocks et al (2007). All scale scores range from 0 to 100.

QLQ-C30 Scoring

For all scales, calculate the raw score (RS) of a scale using the mean of the item scores in the scale as follows:

$$RS = (S1 + S2 + \ldots + Sn) / n$$

where Si: i=1, …, n, are the item scores and n is the number of items with valid scores, assuming the number of items with valid scores meets the minimum requirement as specified in Table 3 or this scale score will be assumed missing.

Use a linear transformation to standardize the raw score in order that scores will range from 0-100:

Global Health Status/QOL = {(RS-1)/range} * 100

Functional Scales = {1- (RS-1)/range} * 100

Symptom Scales = {(RS-1)/range} * 100

where range for each scale is defined in Table 3.

For the Global Health Status/QOL scale and functional scales in QLQ-C30 a higher score represents a better health state and for the symptom scores in QLQ-C30 a lower score represents a better health state.

QLQ-MY20 Scoring

For all scales, calculate the raw score (RS) of a scale using the mean of the item scores in the scale as follows:

$$RS = (S1 + S2 + \ldots + Sn) / n$$

where Si: i=1, …, n, are the item scores and n is the number of items with valid scores, assuming the number of items with valid scores meets the minimum requirement as

specified in Table 3 or this scale score will be assumed missing.

Use a linear transformation to standardize the raw score in order that scores will range from 0-100:

Disease Symptom Scale (DS) = {(RS – 1) / range} * 100

Side Effects of Treatment Scale (SE) = {(RS – 1) / range} * 100

Future Perspective Scale (FP) = {1 - (RS – 1) / range} * 100

Body Image (BI) = {1 - (RS – 1) / range} * 100

where range for each scale is defined in Table 3.

Table 3: QLQ-C30 and QLQ-MY20 Scales and Scoring Details

	Number of Items	Item Range [a]	Item Numbers	Minimum Not Missing
QLQ-C30				
Global Health status/QOL	**2**	6	29,30	1
Functional Scales				
Physical Functioning	5	3	1-5	3
Role Functioning	2	3	6,7	1
Emotional Functioning	4	3	21-24	2
Cognitive Functioning	2	3	20,25	1
Social Functioning	2	3	26,27	1
Symptom Scales / Items				
Fatigue	3	3	10,12,18	2
Nausea/vomiting	2	3	14,15	1
Pain	2	3	9,19	1
Dyspnoea	1	3	8	N/A
Insomnia	1	3	11	N/A
Appetite Loss	1	3	13	N/A
Constipation	1	3	16	N/A
Diarrhoea	1	3	17	N/A
Financial Difficulties	1	3	28	N/A
QLQ-MY20				
Symptom Scales				
Disease Symptoms	6	3	31-36	3
Side Effects of Treatment	10	3	37-46	5
Functional Scales/Items				
Future Perspective	3	3	48-50	1
Body Image	1	3	47	N/A

[a] Range is the difference between the maximum possible value of the Raw Score and the minimum possible value.

For the Disease Symptoms and the Side Effects of Treatment scales a high score represents a high level of symptomatology / problems (symptom scale). For Body Image and Future Perspective scales a high score represents a high level of functioning (functional scale), i.e. a better future perspective or body image.